

WE CLAIM:

1. A method of building a database from text, the method comprising the steps of:
parsing text to identify paths formed by concatenated relationships between words in the text; and
associating, in a computer, paths with each other based on a similarity measure between the paths.
2. The method of claim 1 in which the similarity measure is based on the frequency of occurrence of words in the paths.
3. The method of claim 2 in which the words are at the end points of the paths.
4. The method of claim 1 in which the step of associating paths with each other comprises the step of counting occurrences of words at the end points of specific paths.
5. The method of claim 4 in which the step of associating paths comprises the step of comparing counts of occurrences of words and associating paths based on the counts of occurrences of the words.
6. The method of claim 5 in which paths are associated only when the similarity measure exceeds a threshold.
7. A method of information retrieval, the method comprising the steps of:
initiating a search for electronic information; and
expanding the search by reference to associated paths in a database constructed according to the method of claim 1.
8. The method of claim 7 in which the search is initiated from a location remote from the location of the database.

9. The method of claim 7 in which, in the method steps of claim 1, the similarity measure is based on the frequency of occurrence of words in the paths.
10. The method of claim 9, in which, in the method steps of claim 1, the words are at the end points of the paths.
11. The method of claim 7 in which, in the method steps of claim 1, the step of associating paths with each other comprises the step of counting occurrences of words at the end points of specific paths.
12. The method of claim 11 in which the step of associating paths comprises the step of comparing counts of occurrences of words and associating paths based on the counts of occurrences of the words.
13. The method of claim 12 in which paths are associated only when the similarity measure exceeds a threshold.
14. Computer readable media containing instructions for carrying out the method steps of claim 1.
15. Computer readable media of claim 14 in which, in the instructions for carrying out the method steps of claim 1, the similarity measure is based on the frequency of occurrence of words in the paths.
16. Computer readable media of claim 15 in which, in the instructions for carrying out the method steps of claim 1, the words are at the end points of the paths.
17. Computer readable media of claim 14, in which the instructions for carrying out the step of associating paths with each other comprise instructions for carrying out the step of counting occurrences of words at the end points of specific paths.

18. Computer readable media of claim 17 in which the instructions for carrying out the method step of associating paths comprise instructions for carrying out the method step of comparing counts of occurrences of words and associating paths based on the counts of occurrences of the words.

19. Computer readable media of claim 18 in which, in the instructions for carrying out the method steps, paths are associated only when the similarity measure exceeds a threshold.